

MODELOS DE PREDICCIÓN DE DESERCIÓN DE CLIENTES PARA UNA ADMINISTRADORA DE FONDOS ECUATORIANA CUSTOMER DROPOUT PREDICTION MODELS FOR AN ECUADORIAN FUND MANAGER

María Bohórquez¹, Joyce Torsys², Milton Paredes Aguirre³

Palabras clave:

Clientes
desertores,
Minería de Datos,
Retención de
clientes.

Resumen

La existencia de una empresa está justificada por sus clientes, quienes son considerados como los activos más importantes. Ante mercados más competitivos y donde las necesidades de los clientes son cada vez más exigentes, las empresas buscan eficiencia en el uso y el análisis de datos. Perder clientes es más costoso que atraer nuevos clientes. El estudio sobre el comportamiento del cliente, particularmente su deserción, se ha convertido en una necesidad imperante dentro del ámbito empresarial. En la presente investigación se emplean técnicas de minería de datos para construir modelos de predicción de deserción de clientes, los cuales pueden ser aplicados dentro del mercado de desintermediación financiera. Los modelos estadísticos usados son: Árboles de decisión, bosques aleatorios y regresión logística, estos son evaluados en términos de precisión mediante área debajo de la curva de características de operación del receptor (AUC). La evaluación de los resultados, muestran que el bosque aleatorio tiene un mejor rendimiento que los otros modelos aplicados en el estudio.

Códigos JEL: M31, Y10

Keywords:

Customer Churn,
Data Mining,
Customer
Retention.

Abstract

The existence of a company is justified by its customers, who are active as the most important assets. Faced with more competitive markets and where the needs of customers are increasingly

¹ Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias Sociales y Humanísticas, Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador
E-mail: maleboho@espol.edu.ec

² Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias Sociales y Humanísticas, Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador
E-mail: jtorsys@espol.edu.ec

³ Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias Sociales y Humanísticas, Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador
E-mail: isparede@espol.edu.ec

demanding, companies seek efficiency in the use and analysis of data. Losing customers is more expensive than attracting new customers. The study on customer behavior, specifically attrition, has become a prevailing need within the business environment. In the presentation of research, data mining techniques are used to build models of customer attrition prediction, which can be applied within the financial disintermediation market. The statistical models used are: Decision Trees, Random Forests and Logistic Regression, these are evaluated in terms of accuracy by the area below the receiver operating characteristics curve (ROC). The evaluation of the results, the evaluation that the random forest has a better performance than the other models applied in the study.

JEL Codes: M31, Y10

INTRODUCCIÓN

Conocer el comportamiento futuro de los clientes permite gestionar las relaciones con ellos (CRM) de manera correcta y poder responder a la reducción de costos de la empresa y a la presión competitiva del mercado. Aquellos clientes que dejen de usar los productos de una empresa son denominados desertores, identificarlos permite aplicar estrategias para retenerlos (Guangli , Lingling , Xingsen , & Yong , 2011). La retención de clientes es una estrategia aplicada para evitar que aquellos se desliguen de la empresa y se vayan con competencia. Se alude que, atraer nuevos clientes es más costoso que retener antiguos clientes (Bong-Horng, Ming-Shian, & Cheng, 2007). Por otro lado, incrementar la ratio de retención de clientes un 5% hace que la empresa aumente sus ganancias de un 25% a un 85% (Feinberg & Trotter, 2001). Particularmente, dentro del sector bancario, si se incrementa tan sólo 1% de la ratio de retención de clientes se podría aumentar las ganancias de la banca (Van den Poel & Larivière, 2004).

La deserción de clientes es un problema de alta relevancia que se anida en empresas de diferentes sectores, de hecho, se han levantado estudios dentro áreas como la telefonía móvil (Ahna, Hana, & Lee, 2006) (Bose & Chen, 2009) (Khan, Jamwal, & Sepehri, 2010), redes sociales en línea (Abbasimeh, Setak, & Soroor, 2013), la industria de banca minorista (Klimontowicz, 2014), entre otros. El sector financiero no se escapa de este fenómeno, hay varios estudios

(Dudyala & Ravi, 2008), (Oyeniya & Adeyemo , 2015), que recalcan la necesidad de la industria bancaria en reconectarse con sus clientes. Los bancos no son las únicas entidades clasificadas dentro del sector financiero, también están las empresas pertenecientes al mercado de desintermediación, sin embargo, la literatura no les sugiere solución, particularmente a las administradoras de fondos.

Una de las principales causas que conlleva al problema planteado es la ineficiencia en el análisis de información que se recaudan de los clientes. De ahí radica la importancia de la minería de datos, técnica estadística considerada como una herramienta estratégica importante para las empresas, debido a que, este método permite analizar grandes volúmenes de información desde diferentes perspectivas y sintetizarlos en información valiosa que permite la temprana y efectiva toma de decisiones de las organizaciones empresariales, sirviendo de gran utilidad para crear un perfil de cliente preciso basado en el comportamiento del cliente (Prasad, 2012). De hecho, en los últimos años se ha dado uso a la minería de datos para modelar el comportamiento de los clientes y optimizar la gestión de retención a sus clientes (Kantar, Millward & Brown, 2018).

Por los motivos expuestos, en este estudio se propone un modelo de predicción de deserción de clientes mediante técnica de minería de datos, con el fin de clasificar a los clientes más propensos a desvincularse de la empresa. Los modelos propuestos son árbol de decisión (CART), arboles

aleatorios y regresión logística; estas son las técnicas más usadas por diferentes autores para la predicción de deserción de clientes, los rendimientos de los modelos predictivos se evalúan mediante la Curva Característica de Funcionamiento (AUC) para sugerir un modelo estadístico adecuado.

MARCO TEÓRICO

Dentro de la academia y el sector empresarial se ha investigado sobre las razones por las cuales un cliente permanece en una empresa; según investigaciones de marketing la empresa debe crear una relación a largo plazo donde se basa en la satisfacción, confianza, calidad o el valor percibido por el cliente (Polo Redondo & Sesé Olivan, 2009); en cambio investigaciones aplicando estrategias según el Customer Relationship Management (CRM) los clientes tienen una lealtad cuando perciben que la empresa les brinda un servicio o bien que no lo ofrece la competencia (Bolton, Kannan, & Bramlett, 2000). En los últimos años el uso se ha dado importancia al uso de la Data Mining como una herramienta con mayor poder estadístico que permita modelar el comportamiento de sus clientes así optimizar las oportunidades de retener a sus clientes (Gonçalves, 2018).

Retención de clientes

La retención de clientes consiste en mantener la relación comercial construida entre un proveedor y un cliente (Gerpott, Rams, & Schi, 2001). La retención de clientes es relevante para una empresa por la aplicación de CRM ya que fortalece la lealtad de los clientes evitando incurrir en costos algo para la captación de nuevos clientes. (Kim, Jung, Suh, & Hwang, 2006); La retención de clientes es el primer método eficaz para la supervivencia y el crecimiento de instituciones financieras como la banca, de hecho, está por encima de otros factores claves como el enfoque en tecnología, el enfoque en segmentos específicos del mercado y el enfoque en productividad y eficiencia (Chitra & Subashini, 2011).

Clientes desertores

La definición de clientes desertores varía según el campo de aplicación al que se realiza el estudio. En la industria de telecomunicaciones, un cliente desertor puede ser un término usado para determinar la rotación de clientes de una empresa a otra empresa competidora (Junxiang, 2002). En el caso de compañías de internet, también puede ser definido como la propensión de los clientes a dejar de hacer negocios con la empresa en un determinado periodo de tiempo (Guangli, Lingling, Xingsen, & Yong, 2011). Dentro del sector bancario, un cliente desertor es aquel que cierra todas sus cuentas bancarias y cesa los negocios con el banco en estudio (Chitra & Subashini, 2011).

Modelos de predicción de clientes desertores aplicando minería de datos

En un mundo globalizado y competitivo la aplicación de minería de datos para el manejo de grandes volúmenes de información se ha convertido en una herramienta importante para mejorar la toma de decisiones. La minería de datos mantiene procesos y algoritmos que permiten a las empresas extraer conocimiento relevante que está oculto dentro de los datos corporativos, con el fin de extender y mejorar la comprensión del negocio, mediante la utilización de sofisticados algoritmos de búsqueda de datos estadísticos (Oyeniya & Adeyemo, 2015).

Se han realizado varios estudios sobre modelos la deserción futura de los clientes mediante el uso de minería de datos, la finalidad de estos es identificar las señales de abandono temprano y reconocer a los usuarios que tienen una alta propensión de irse voluntariamente. Las técnicas de minería de datos que han sido aplicadas en este tipo de modelos son: Árboles de decisión, redes neuronales, reglas de asociación, regresión logística, árboles aleatorios y máquinas de vectores de soporte (SVM). Tales investigaciones se han realizado en diferentes sectores como: Telecomunicación, comercial minorista, bancario, entre otros. A continuación, se presenta una breve revisión literaria de estudios realizados dentro del sector financiero.

TABLA 1
Revisión de Literatura

Autor	Industria	Metodología	Datos
Wouter Buckinx & DirkVan den Poel (2005)	Bienes de consumo minorista	Regresión logística, Redes neuronales y bosques aleatorios	Se usó registro de 158884 clientes que presentan movimientos de compras desde abril del 2000 a enero del 2001.
Teemu Mutanen (2006)	Banca	Regresión Logística	Datos de clientes de un banco finlandés. Los datos disponibles se recopilaban desde el período de diciembre de 2001 hasta septiembre de 2005 (115000 datos)
Dudyala Anil Kumar and V. Ravi (2008)	Banca	Modelos de MLP, LR, árbol de decisión, RF, RBF, SVM y SMOTE. La clasificación de CART.	14814 datos de un banco latinoamericano, 13812 no desertores y 1002 desertores.
Yaya Xie et al. (2009)	Banca	Bosques aleatorios	Datos de registros transaccionales de más de 20000 clientes, tiempo de recolección no especificado.
Guangli Nie et al. (2011)	Tarjeta de créditos/Banca	Regresión logística y árbol de decisión	4997 cuentas de banco comercial de china en el año 2006.
K. Chitra & B.Subashini (2011)	Banca	CART	Consideraron el comportamiento de los últimos 3 meses de transacción de los clientes activos del banco en estudio, teniendo en cuenta 1000 registros.
U. Devi Prasad & S. Madhavi (2012)	Banca	Árbol de decisión CART y C-5	1484 el número de clientes de Banco Nacional de India en un periodo de 3 meses; 1163 cuentas activas y 311 inactivas
Soeini & Rodpysh (2012)	Seguros	Árbol de decisión CART, método K-means	Recolección de información mediante cuestionario y para ese propósito se utilizó variables de modelado de abandono de acuerdo con entrevistas primarias de reporteros de la industria de seguros.
Md. Rafiqul islam and Md. Ahsan Habib (2015)	Banca comercial minorista	Árbol de decisión podado	Datos de registros transaccionales de clientes, durante un periodo de obtuvieron 18 meses.
Davoud Gholamiangonaba et al. (2019)	Banca	Regresión General Red Neuronal, Naive Bayes, MLP, SVM, RBF NN)	Los datos utilizados en la investigación son de un banco iraní, 860 datos entre febrero de 2013 y junio de 2013.

METODOLOGÍA

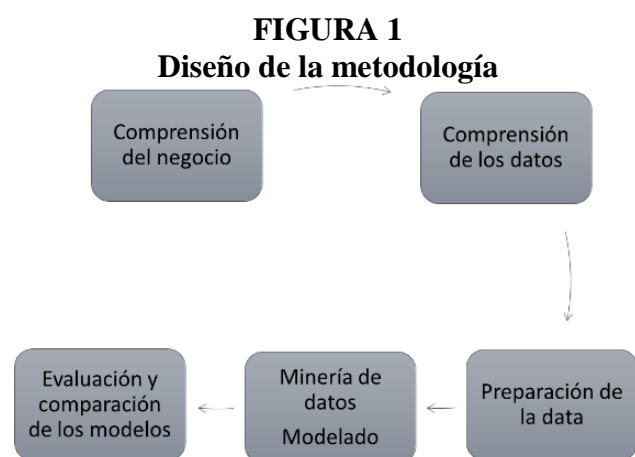
En esta sección se detalla el proceso a seguir para cumplir cada uno de los objetivos del estudio. Para la identificación de clientes desertores se usaron dos tipos de investigación cualitativa y cuantitativa. Dentro de la

investigación cualitativa se aplicaron entrevistas a profundidad, la ventaja de esta técnica radica en los beneficios de crear un ambiente libre donde el dialogo fluye y se aborda varios temas (Rodrigues, Hoffmann, Mackedanz, & Hoffmann, 2011). Se entrevistaron a los responsables de las áreas de administradora de

cuentas, marketing y de innovación, con el fin de comprender mejor el negocio.

La parte cuantitativa de este estudio parte de la preparación de la data, luego se ejecutan los modelos de predicción usando técnicas de minería de datos y finalmente se evalúa el rendimiento de los modelos. Para la identificación de posibles clientes desertores se eligió el modelo que tenga el mejor rendimiento con respecto a la predicción de los datos, se evaluó por medio de la Curva Característica de Funcionamiento (AUC). Así mismo, el programa estadístico en el cual se usó para el proceso de análisis cuantitativo fue R Studio.

A continuación, se ilustra en un diagrama la metodología usada en el proyecto.



Minería de datos

La técnica de minería de datos consiste en la selección, limpieza, transformación y reducción de la data, además de la interpretación, evaluación y desarrollo de modelos para el apoyo de decisiones, teniendo como objetivo extraer información útil de los datos (Kantardzie, 2003). Por tal motivo, las técnicas de minería de datos contienen un potencial de poder predictivo elevado, debido a su capacidad de analizar grandes volúmenes de información y su hábil búsqueda de patrones y relaciones sistemáticas a través de métodos de aprendizaje, inteligencia artificial y estadística.

Descripción de los datos

La base de datos usada en este estudio contiene 105353 registros de clientes pertenecientes a la AFE, donde se detallan atributos sociodemográficos del cotizante y del asesor, también contiene variables de comportamiento del cliente, y variables que describen la empresa donde trabaja el cotizante, dando un total de 90 atributos. Sin embargo, se eliminaron atributos como el nombre del cliente, del asesor, del oficial de cuentas, fecha de nacimiento, fecha de ingreso, código del asesor, entre otras variables que no afectan a la predicción de deserción del cliente, quedando con un total de 20 atributos. La literatura sugiere usar variables sociodemográficas y de comportamiento del cliente para la ejecución del modelo de predicción. Para comprobar la utilidad de estas variables se usaron estadísticas descriptivas y además la experiencia de expertos.

Finalmente se agruparon los siguientes grupos de variables:

Variables sociodemográficas del cliente: Estos atributos caracterizan al cliente según su edad, género, provincia e ingreso percibido.

Variables de comportamiento del cliente: Estos atributos capturan el comportamiento transaccional del cliente según el monto de depósito, el tipo de depósito, el saldo actual del fondo, el sistema de aporte, las transacciones de los últimos 12 meses.

Preparación de los datos

Siguiendo artículos referenciales como (Oyeniyi et al., 2015; Chih-Fong et al., 2010; Kaur et al., 2013), el tratamiento de la data se basó en la eliminación de valores faltantes, transformación y estandarización de las variables y el balanceo del número de observaciones y la definición del periodo de estudio

Valores faltantes

Esta parte del tratamiento consiste identificar aquellas variables con muchos valores ausentes, luego según sea conveniente se procede a

eliminarlas o a llenar esos casilleros vacíos con promedios o modas.

Transformación de las variables

En la base de datos hay variables categóricas como tipo de monto, parroquia, entre otras, por tal motivo, tales variables se las transforman a números.

Estandarización de las variables

Con el fin de mejorar el funcionamiento de los algoritmos usados en este estudio, se empleó la técnica de estandarización por el método de MinMax el cual fue recomendado por expertos en el análisis de los datos de AFE dado que facilita la comparación de variables con distintas unidades de medida.

Balanceo de datos

Dentro de la data, el porcentaje de clientes inactivos asciende al 16% y el 84% son clientes activos, claramente existe un problema de datos desbalanceados. Pocos estudios sobre deserción de cliente le dan importancia a la identificación de data desbalanceada, según la literatura revisada, la técnica de balanceo de datos más usada es SMOTE, esta consiste en un sobre muestreo sin remplazo a la clase minorista (clientes inactivos), creando muestras sintéticas, de tal manera que se logre ampliar la región de datos que corresponde a muestras minoristas (Chawla et al, 2004).

Periodo de estudio seleccionado

La base de datos captura movimiento transaccional desde octubre del 2018 hasta octubre del 2019, para este tipo de estudios generalmente se analiza el comportamiento del cliente dentro de 12, 6 o 3 meses. El periodo de estudio que se seleccionó fue de tres meses, debido a que, hay clientes que desertaron desde octubre, entonces trabajando bajo el supuesto de que el 50% de la data hayan abandonado la empresa desde octubre, entonces el modelo no podrá captar el comportamiento del cliente antes de la rotación de manera adecuada, esta situación de la data se asemeja a (Prasad & Madhavi, 2012), tomando la solución que ellos optaron, se

procedió a elegir los últimos tres meses de actividad antes de que el cliente se retire, es decir, que el periodo de tiempo difiere por cotizante.

Datos de entrenamiento y datos de prueba

Tomando en cuenta estudios de (Kaur, Singh, & Sharma, 2013; Oyeniyi & Adeyemo, 2015; Olle & Cai, 2014) la data de entrenamiento corresponde a un 70% y el 30% a la data de prueba.

Modelos de predicción

Árbol de decisión

La técnica de árbol de decisión es un método no paramétrico, lo cual permite saltarse supuestos distribucionales, además es capaz de detectar interacciones y modelar relaciones no lineales. Esta técnica consiste en otorgar resultados de distintas combinaciones de decisiones y eventos mediante particiones recursivas que usan reglas de clasificación. El algoritmo, Árbol de Clasificación y Regresión (CART), genera un árbol de decisión binario según una función de atributo único, utilizando el índice de Gini para determinar la mejor división, este método es muy recomendado por (Chitra et al., 2011; Prasad & Madhavi et al., 2012; Guangli et al., 2006; Kaur et al., 2013), debido a que, CART no usa regla de detención, es decir el árbol crece y luego se poda, garantizando que no se pasen por alto patrones importantes al detenerse demasiado pronto.

Árboles Aleatorios

Es un método ensamblador cuyo fin es buscar el modelo de predicción óptimo de un conjunto de grupo de clasificadores (árboles de decisión), un modelo que mantenga un balance entre su error de predicción y su varianza. La ventaja de este método es la creación de árboles de forma aleatorios usando los conjuntos de entrenamiento, como también las valoraciones que da los árboles a cada una de las variables (Buckinx & Van den Poel, 2005). Como clasificar base el modelo realiza en cada nodo se selecciona un subconjunto aleatorio dando así la mejor selección de división para ese nodo (Burez & Van den Poel, 2009).

Regresión logística

Es un tipo de algoritmo estadístico, cuyo objetivo es modelar la probabilidad de un suceso de una variable dependiente en función de otras variables explicativas o predictores, cabe recalcar que para la construcción de este modelo tampoco se realizan supuestos sobre la distribución de probabilidad de las variables. La regresión logística es considerada por (Shyug Lee et al., 2004) como un poderoso algoritmo, debido a que, se puede identificar la significancia de las variables explicativas, por tal motivo autores como (Guangli et al., 2006; Van den Poel et al., 2004), han optado por usar es técnica de minería de datos para sus estudios sobre deserción del cliente.

Evaluación de los modelos

En esta sección se procede a identificar el modelo con mejor acierto en la predicción de

deserción de cliente. Debido a las referencias de (Kumar et al., 2008; Islam et al., 2015), para evaluar los modelos propuestos se usa el método de área bajo la curva característica de funcionamiento del receptor (AUC-ROC). (AUC-ROC) es un gráfico donde en el eje “y” se sitúan los puntos de verdaderos positivos (sensibilidad) y en el eje “x” los falsos positivos (1-especificidad), con esto mencionado, cuanto más se aproxime una curva ROC hacia arriba y hacia la izquierda del plano, entonces más alta es la exactitud del modelo.

RESULTADOS

Variables predictoras seleccionadas

Luego de la reducción de aquellas variables predictoras que estén altamente correlacionadas y cuyas varianzas sean cercanas “0”, se terminó escogiendo a 10 predictores, los cuales son detallados en la tabla 2.

TABLA 2
Variable predictoras seleccionadas

Variable	Tipo de dato	Descripción
“Edad”	Numérico	Edad del cliente durante el periodo de estudio.
“Ingreso”	Numérico	Ingreso monetario del cliente durante el periodo de estudio.
“Prommonto”	Numérico	Promedio mensual de los últimos tres meses (varía según cotizante).
“Prima”	Numérico	Prima del cliente durante el periodo de estudio.
“Saldoactual”	Numérico	Saldo del fondo horizonte del cliente durante el periodo de estudio.
“Empresa”	Independiente, Relación de dependencia (Nominal)	Si el cliente trabaja en relación de dependencia o no.
“Categoríapempresa”	EE, NOEE (Nominal)	Si la empresa es estratégica o no.
“Provinciacliente”	Nominal	Provincia en la que el cliente reside.
“Genero”	Femenino, Masculino (Nominal)	Género del cliente.
“Sistemadeaporte”	TCredito, Rol, Individual, Cuentacorrente, Cuentaahorro (Nominal)	Forma de pago del cliente, ya sea, mediante débito de Cuenta de ahorro, corriente, rol, tarjeta de crédito o depósito de manera individual.

Predicción de clientes desertores ejecutando un árbol de decisión (algoritmo CART)

De acuerdo con las reglas de la tabla 2 se puede observar que las variables más importantes para entender el comportamiento de clientes desertores

son el saldo del fondo que mantienen, el sistema de aporte, empresa y categoría de la empresa.

Si el saldo del fondo horizonte del cliente es menor a \$65.49, entonces es probable que deserten. Por otro lado, si tiene clientes que

mantienen un saldo mayor o igual a \$65.49 y menor a \$870.66 y la forma de pago la hacen de manera individual, o descuento mediante rol o de la tarjeta de crédito, entonces es probable que sus clientes se vayan de la empresa. Si existen clientes que mantienen un saldo mayor o igual a \$65.49 y menor a \$143.83, la forma de pago la hace mediante débito de su cuenta corriente o de

ahorro, y no trabajan en relación de dependencia, entonces es probable que deserten. Finalmente, si hay clientes que mantienen un saldo mayor o igual a \$65.49 o a \$143.83, la forma de pago la hace mediante débito de su cuenta corriente o de ahorro, no trabajan en relación de dependencia, y la categoría de la empresa es estratégica, entonces es probable que deserten

TABLA 3
Resultado de CART

Regla de decisión	Clase predicha	Número de casos
Saldoactual < \$65.49	Desertores	10176
Saldoactual ≥ \$65.49, sistemadeaporte= Individual, Tcredito y Rol, y saldoactual < \$870.66	Desertores	2095
Saldoactual ≥ \$65.49, sistemadeaporte= Cuentacorriente y Cuentaahorro, saldoactual < \$143.83, y empresa=Independientes	Desertores	599
Saldoactual ≥ \$65.49, sistemadeaporte= Cuentacorriente y Cuentadeahorro, saldoactual ≥ \$143.83, empresa=Independientes, y categoriaempresa= Empresa estratégica	Desertores	299

Predicción de clientes desertores ejecutando un bosque aleatorio (Random Forest)

De acuerdo con los resultados del modelo estadístico, los atributos más importantes para la predicción de clientes desertores son el saldo del fondo, el tipo de sistema de aporte, y si el cliente trabaja en relación de dependencia o no. A continuación, se detallan los resultados en la tabla 4.

TABLA 4
Resultados de Bosque Aleatorio

Nombre de las variables	Rango	Clase predicha
Saldoactual	Menor a \$80	Desertor
Sistemadeaport e	Individual, Rol, Tarjeta de crédito	Desertor
Empresa	Independientes	Desertor

Se aplicó el efecto marginal de los predictores sobre la variable de resultado para identificar los rangos de clasificación de cada variable. Los resultados muestran que, si el cliente mantiene un

saldo del fondo menor a \$80 tiene un mayor efecto en la respuesta desertora del cliente. Si el cliente mantiene una forma de pago de manera individual, descuento por Rol, o por tarjeta de crédito, es más probable que deserten. Si trabaja de manera independiente entonces es más probable que deserte. La tabla 5 resume lo expuesto.

TABLA 5
Efecto marginal de los predictores sobre variable de resultado

Orden	Nombre de las variables	Disminución media de Gini
1	“Saldoactual”	6851.27
2	“Sistemadeaporte”	1250.04
3	“Empresa”	1015.62

Predicción de clientes desertores ejecutando una regresión logística

La tabla 6 muestra los resultados de la regresión logística.

TABLA 6
Regresión logística

Pr(> z)	Estimado	Std.	Error z	Valor
Intercepto	4.44	0.11	40.23	< 2e-16 ***
Edad	-1.03	0.14	-7.36	1.84e-13 ***
Prom.monto	32.57	2.15	05.15	0.0001
Ingreso	-1.8	0.22	-8.11	5.05e-16 ***
Saldoactual	-837.97	12.88	-65.07	< 2e-16 ***
Empresa	-1.22	0.05	-26.12	< 2e-16 ***
Cat.empresa	-0.64	0.04	-15.58	< 2e-16 ***
Cta de ahorro	-1.74	0.08	-22.93	< 2e-16 ***
Cta. corriente	-0.24	0.10	-2.51	0.0121 *
Apor. Ind.	2.60	0.22	11.89	< 2e-16 ***
Tarj. crédito	0.25	0.11	2.20	0.0276 *
Género	-0.26	0.03	-7.77	7.68e-15 ***

Los resultados de la regresión logística muestran que, mientras mayor sea el ingreso, la edad, y el saldo del fondo pues menor será la probabilidad que el cliente deserte. Si el cliente maneja un tipo de pago mediante tarjeta de crédito y de manera individual, entonces es más probable que deserte. Si sus clientes son mujeres entonces es más probable que deserten. Si trabajan en relación de dependencia entonces es más probable que no deserten. El promedio de monto de los últimos 3 meses no es significativo.

Evaluación de los modelos

Los tres modelos fueron evaluados mediante la Curva Característica de Funcionamiento, los resultados se muestran en la tabla 7.

TABLA 7
Resultados Curva ROC-AUC

Modelos	AUC
Árbol de decisión CART	88%
Bosque aleatoria	93%
Regresión logística	91%

Según los resultados del área bajo la Curva ROC, el bosque aleatorio obtiene un 93% de predicción correctas, superando a los modelos de regresión logística y al árbol de decisión, por lo tanto, el mejor modelo de predicción de deserción

de clientes para este estudio es el de bosques aleatorios.

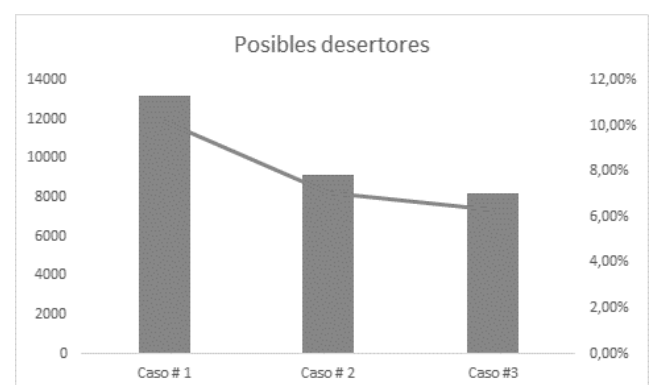
Cientes activos en riesgo de abandono

El bosque aleatorio para perfilar al cliente desertor arroja como variables importantes el saldo actual, empresa y sistema de aporte. Al analizar la data bruta del estudio (150000 cotizantes activos o no) se identifica que, del total de clientes activos, el 10.16% poseen un saldo menor a \$80, el 7.04% mantiene una forma de pago mediante descuento por rol, tarjeta de crédito o aporte individual, y el 6.32 % no trabajan en relación de dependencia. En la tabla 8 se detalla lo comentado.

TABLA 8
Predicción de Posibles Desertores

Casos	Número de casos	Porcentaje
1	13202	10,16%
2	9157	7,04%
3	8218	6,32%
Total	30577	23,52%

FIGURA 2
Distribución de estado y saldo actual



El 23.52% de los cotizantes activos poseen un perfil desertor, donde mayor peso tienen aquellos que mantienen un saldo menor a \$80, le sigue el sistema de aporte y luego su tipo laboral.

CONCLUSIONES

El estudio del comportamiento del cliente es un insumo indispensable para las empresas, especialmente su deserción, el alto volumen de

información generado en el mercado y el uso de estadísticas débiles impide a muchas empresas capturar de manera precisa las señales de deserción, por lo tanto, se optó por la minería de datos como solución, siendo esta una herramienta que permite optimizar la gestión y análisis de datos.

Dentro de la academia, el estudio de modelos de predicción de clientes desertores usando minería de datos se ha realizado principalmente en bancos, sin embargo, no se lo ha hecho en una administradora de fondos. Este estudio presenta una solución a un problema específico de una empresa en particular, en este caso la AFE. Sin embargo, la perfilación de clientes desertores mediante las técnicas de minería de datos usadas en la presente investigación es útil para cualquier empresa perteneciente al mercado de desintermediación financiera que quiera clasificar a sus clientes desertores y direccionar de mejor manera sus estrategias de retención y fidelización.

Los modelos estadísticos aplicados permiten identificar quienes son los clientes que pueden irse de la empresa, siendo el modelo de bosques aleatorios el óptimo. Las variables más relevantes son el saldo de la cuenta, tipo de sistema de aportación y si trabaja de forma independiente. Según la perfilación hecha en el estudio de los clientes activos que tiene la empresa, se determinó que 30577 afiliados tienen características de abandono.

REFERENCIAS

- Abbasimeh, H., Setak, M., & Soroor, J. (2013). A framework for identification of high-value customers by including social network based variables for churn prediction using neuro-fuzzy techniques. *International Journal of Production Research*, 51(4), 1279-1294.
- Ahna, J.-H., Hana, S.-P., & Lee, Y.-S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications Policy*, 30(10), 552-568.
- Bolton, R., Kannan, P., & Bramlett, M. (2000). Implications of Loyalty Program Membership and Service Experiences for Customer Retention and Value. *Journal of the Academy of Marketing Science*, 95-108.
- Bong-Horng, C., Ming-Shian, T., & Cheng, S. (2007). Toward a hybrid data mining model for customer retention. *Knowledge-Based Systems*, 703-718.
- Bose, I., & Chen, X. (2009). Hybrid Models Using Unsupervised Clustering for Prediction of Customer Churn. *Journal of Organizational Computing & Electronic Commerce*, 19(2), 133-151.
- Buckinx, W., & Van den Poel, D. (2005). Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, 252-268.
- Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 4626-4636.
- Chawla, N., Bowyer, K., L.O., H., & Kegelm, W. (2004). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 324-357.
- Chih-Fong, T., & Mao-Yuan, C. (2010). Variable selection by association rules for customer churn prediction of multimedia on demand. *Expert Systems with Applications*, 2006-2015.
- Chitra, K., & Subashini, B. (2011). Customer Retention in Banking Sector using Predictive Data Mining Technique. *The 5th International Conference on Information Technology*.
- Dudyala, A. K., & Ravi, V. (2008). Predicting credit card customer churn in banks using data mining. *Institute for Development and Research in Banking Technology*, 1(1), 4-28.
- Feinberg, R. A., & Trotter, M. (2001). Immaculate deception: the unintended negative effects of the CRM revolution: maybe we would be better off without customer relations management. *Defying the limits*, 26-31.
- Gonçalves, D. (2018). *El Reto de la Fidelización en las Empresas Españolas*. Madrid.
- Guangli, N., Lingling, Z., Xingsen, L., & Yong, S. (2011). The Analysis on the Customers Churn of Charge Email Based on Data Mining Take One

Internet Company for Example. *Institute of Electrical and Electronics Engineers*, 843-847.

Islam, R., & Habib, A. (2015). A Data Mining Approach to Predict Prospective Business Sectors for Lending in Retail Banking Using Decision Tree. *International Journal of Data Mining & Knowledge Management Process*, 13-23.

Kantardzie, M. (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*. New York: Wiley-IEEE Press.

Kaur, M., Singh, K., & Sharma, N. (2013). Data Mining as a tool to Predict the Churn Behaviour among Indian bank customers. *International Journal on Recent and Innovation Trends in Computing and Communication*, 720-725.

Khan, A., Jamwal, S., & Sepehri, M. (2010). Applying Data Mining to Customer Churn Prediction in an Internet Service Provider. *Applying Data Mining to Customer Churn Prediction in an Internet Service Provider*, 9(7), 8-14.

Kim, S.-Y., Jung, T.-S., Suh, E.-H., & Hwang, H.-S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications*, 101-107.

Klimontowicz, M. (2014). Customer-Centricity Evolution as a Foundation of Bank's Competitive Strategy. *Journal of Economics and Management*, 16.

Kumar, D., & Ravi, V. (2008). Predicting credit card customer churn in banks using data mining. *Int. J. Data Analysis Techniques and Strategies*, 4-29.

Junxiang, L. (2002). Predicting Customer Churn in the Telecommunications Industry An Application of Survival Analysis Modeling Using SAS. *Sprint Communications Company*.

Olle, G. D., & Cai, S. (2014). A Hybrid Churn Prediction Model in Mobile. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 55-62.

Oyeniyi, A., & Adeyemo, A. (2015). Customer Churn Analysis in Banking Sector Using Data Mining Techniques. *African Journal of Computing & ICT*, 165-174.

Polo Redondo, Y., & Sesé Olivan, F. J. (2009). La retención de los clientes un estudio empírico de sus determinantes. *Revista Española de Investigación de Marketing*, 117-137